A lossless data compression method enabling fast decompression

Abstract of Disclosure

A lossless data compression method that enables fast decompression is presented. To roughly characterize the method essential is that the finding of the repetitions is based on symbol pairs. The method repeatedly finds pairs formed by two adjacent symbols and replaces occurrences of common pairs with one new symbol. Similar prior art methods store the replacement rules into a separate entity that itself has to be compressed using some technique. The presented method does not need the replacement rules – the compressed symbol sequence is built in a new way, which enables that the replacement rules can be inferred from the symbol sequence. In prior art decompression, the replacement rules are decompressed and then consulted to interpret the symbol sequence and produce the original uncompressed input. The presented method enables decompression without recursion in one sequential traversal of the compressed input. To overcome the need to have a separate list of replacement rules the following method is used: The first one of the occurrences of a pair is not replaced. Later in the algorithm when no more symbols are replaced every introduced new symbol that still exists is turned to be a pointer to the corresponding first occurrence of the pair. A pointer is the position of the first symbol of a pair. To distinguish between atomic symbols i.e. symbols from original uncompressed input and pointers a constant value greater than any atomic value is added to a pointer value. The decompression traverses the compressed sequence and keeps track of each compressed position's decompressed length and starting position in the decompressed sequence. For every pointer position in compressed input, the number of decompressed symbols indicated by the pointed position and the next one are copied to decompressed output.

Figures